1. A method comprising:

(a) providing at least a first and a second independent discovery data set wherein:

(i) the data sets comprise a plurality of forms of biological state classes;

(ii) each data set comprises a plurality of data points, wherein each data point exhibits one form of a biological state class and each data set comprises a plurality of data points belonging to each of the classes;

(iii) each data point comprises a plurality of data elements, each data element characterized by a value, wherein all data points share a plurality of common data elements; and

(b) qualifying each common data element, independently for each dataset, based on the ability of the data element to classify a data point into a form of biological state class, as a function of data element value;

(c) selecting an initial subset of data elements within each data set, and

(d) selecting an intersection subset of data elements from the initial subsets, wherein each data element in the intersection subset is a member of a majority of the initial subsets.

2. The method of claim 1, wherein the step of selecting the initial subsets comprises using the discovery data sets to train a learning algorithm wherein the learning algorithm ranks the data elements based on a quantitative measure of ability to classify.

3. The method of claim 2, wherein the learning algorithm is a supervised learning algorithm.

4. The method of claim 2, wherein the learning algorithm is an unsupervised learning algorithm.

5. The method of claim 3, wherein the training comprises using support vector machine analysis.

6. The method of claim 2, wherein the training comprises performing linear discrimination analysis.

7. The method of claim 2, wherein, the training comprises performing unified maximum separability analysis (UMSA).

8. The method of claim 1, further comprising independently re-sampling data elements in each data set.

9. The method of claim 1, further comprising, selecting candidate biomarkers from selected data elements and testing one or more of the candidate biomarkers on a validation data set.

10. The method of claim 1, wherein the biological state class comprises a cell state.

11. The method of claim 1, wherein the biological state class is a patient status.

12. The method of claim 1, wherein the biological state class is selected from the group consisting of: presence of a disease; absence of a disease; progression of a disease; risk for a disease; stage of disease; likelihood of recurrence of disease; a genotype; a phenotype; exposure to an agent or condition; a demographic characteristic; resistance to agent, sensitivity to an agent, and combinations thereof.

13. The method of claim 12, wherein the genotype is selected from the group consisting of an HLA haplotype; a mutation in a gene; a modification of a gene, and combinations thereof.

14. The method of claim 12, wherein the agent is selected from the group consisting of a toxic substance, a potentially toxic substance, an environmental pollutant, a candidate drug, and a known drug.

15. The method according to claim 12, wherein the demographic characteristic is selected from the group consisting of: age, gender, weight; family history; and history of preexisting conditions.

16. The method according to claim 12, wherein sensitivity to an agent comprises responsiveness to a drug.

17. The method of claim 9, wherein the one or more candidate biomarkers are diagnostic of the presence of a disease, risk of developing a disease, risk of recurrence of a disease, or stage of the disease.

18. The method of claim 1, wherein values of the data elements in a data point represent levels and/or frequency of components in a data point sample.

19. The method of claim 18, wherein components are selected from the group consisting of: nucleic acids, proteins, polypeptides, peptides, carbohydrates and modified or processed forms thereof.

20. The method of claim 18, wherein levels of components are measured by an expression profiling assay.

21. The method of claim 20, wherein the expression profiling assay comprises measuring the amount and/or form of a nucleic acid.

22. The method of claim 21, wherein expression profiling comprises measuring amplification, mutation, and/or modification of DNA.

23. The method of claim 20, wherein the expression profiling assay comprises measuring the amount and/or form of a protein, polypeptide or peptide.

24. The method of claim 23, wherein the expression profiling assay comprises mass spectrometry.

25. The method of claim 24, wherein the expression profiling assay comprises SELDI analysis.

26. The method of claim 20, wherein the expression profiling assay comprises measuring the amount and/or form of a carbohydrate.

27. The method of claim 1, wherein data elements of data points comprise data relating to the cellular localization of components in a sample.

28. The method of claim 20, wherein expression profiling comprises:

(a) contacting samples with a substrate comprising binding partners for specifically binding to sample components having selected characteristics and

(b) identifying sample components bound to the substrate.

29. The method according to claim 28, wherein binding partners are selected from the group consisting of cationic molecules; anionic molecules; metal chelates; antibodies; single- or double-stranded nucleic acids; proteins, peptides, amino acids; carbohydrates; lipopolysaccharides; sugar amino acid hybrids; molecules from phage display libraries; biotin; avidin; streptavidin; and combinations thereof.

30. The method of claim 28, wherein the binding partners are arrayed on the substrate.

31. The method of claim 2, wherein an assay used to measure levels of data elements in training data sets from which candidate biomarkers are identified is different from an assay used to measure data elements in a validation data set used to validate the candidate biomarker.

32. The method of claim 31, wherein the assay used to measure levels of data elements in training data sets is SELDI.

33. The method of claim 31 or 32, wherein the assay used to measure levels of data elements in validation data sets is an immunoassay.

34. The method of claim 1, wherein the independent discovery data sets are collected from different locations, using different collection protocols, and/or are collected from different populations.

35. The method of claim 1, wherein each discovery data set is from a different clinical trial site.

36. A computer program product comprising a computer readable medium having:

(a) a first computer readable program code providing instructions for causing a computer to input data relating to at least first and second independent discovery data sets wherein:

i) the data sets comprise a plurality of forms of biological state classes;

ii) each data set comprises a plurality of data points, wherein each data point exhibits one form of a biological state class and each data set comprises a plurality of data points belonging to each of the classes; and

iii) each data point comprises a plurality of data elements, each data element characterized by a value, wherein all data points share a plurality of common data elements;

(a) a second computer readable program code providing instructions for qualifying each common data element, independently for each data set, based on the ability of the data element to classify a data point into a biological state class, as a function of data element value and for selecting an initial subset of data elements within each data set, and

(b) a third computer readable program code providing instructions for selecting an intersection subset of data

elements from the initial subsets, wherein each data element in the intersection subset is a member of a majority of the initial subsets.

37. The computer program product according to claim 36, wherein selecting the initial subsets comprises using the discovery data sets to train a learning algorithm wherein the learning algorithm ranks the data elements based on a quantitative measure of ability to classify.

38. The computer program product according to claim 37, wherein the learning algorithm is a supervised learning algorithm.

39. The computer program product according to claim 37, wherein the learning algorithm is an unsupervised learning algorithm.

40. The computer program product of claim 37, wherein training comprises support vector machine analysis.

41. The computer, program product of claim 37, wherein training comprises linear discrimination analysis.

42. The computer program product of claim 37, wherein training comprises combining support vector machine analysis and linear discrimination analysis.

43. The computer program product of claim 37, wherein training comprises performing unified maximum separability analysis (UMSA).

44. The computer program product of claim 36, further comprising program code for independently re-sampling data elements in each data set.

45. The computer program product of claim 37, further comprising program code for selecting candidate biomarkers based on ranking by the learning algorithm

and for testing one or more of the candidate biomarkers on a validation data set.

46. The computer program product of claim 36, wherein the biological state class comprises a cell state.

47. The computer program product of claim 36, wherein the biological state class comprises a patient status.

48. The computer program product of claim 36, wherein the biological state class is selected from the group consisting of: presence of a disease; absence of a disease; progression of a disease; risk for a disease; stage of disease; likelihood of recurrence of disease; a genotype; a phenotype; exposure to an agent or condition; a demographic characteristic; resistance to agent, sensitivity to an agent, and combinations thereof.

49. The computer program product of claim 48, wherein the genotype is selected from the group consisting of an HLA haplotype; a mutation in a gene; a modification of a gene, and combinations thereof.

50. The computer program product of claim 48, wherein the agent is selected from the group consisting of a toxic substance, a potentially toxic substance, an environmental pollutant, a candidate drug, and a known drug.

51. The computer program product of claim 48, wherein the demographic characteristic is selected from the group consisting of: age, gender, weight; family history; and history of preexisting conditions.

52. The computer program product of claim 48, wherein sensitivity to an agent comprises responsiveness to a drug.

53. The computer program product of claim 45, wherein the one or more candidate biomarkers are diagnostic of the presence of a disease, risk of developing a disease, risk of recurrence of a disease, or stage of the disease.

54. The computer program product of claim 36, wherein values of the data elements in a data point represent levels and/or frequency of components in a data point sample.

55. The computer program product of claim 54, wherein components are selected from the group consisting of: nucleic acids, proteins, polypeptides, peptides, carbohydrates and modified or processed forms thereof.

56. The computer program product of claim 54, wherein levels of components are measured by an expression profiling assay.

57. The computer program product of claim 56, wherein the expression profiling assay comprises measuring the amount and/or form of a nucleic acid.

58. The computer program product of claim 56, wherein expression profiling comprises measuring amplification, mutation, and/or modification of DNA.

59. The computer program product of claim 56, wherein the expression profiling assay comprises measuring the amount and/or form of a protein, polypeptide or peptide.

60. The computer program product of claim 56, wherein the expression profiling assay comprises mass spectrometry.

61. The computer program product of claim 56, wherein the expression profiling assay comprises SELDI analysis.

62. The computer program product of claim 56, wherein the expression profiling assay comprises measuring the amount and/or form of a carbohydrate.

63. The computer program product of claim 36, wherein data elements of data points comprise data relating to the cellular localization of components in a sample.

64. The computer program product of claim 56, wherein expression profiling comprises:

(a) contacting samples with a substrate comprising binding partners for specifically binding to sample components having selected characteristics; and

(b) identifying sample components bound to the substrate.

65. The computer program product of claim 64, wherein binding partners are selected from the group consisting of cationic molecules; anionic molecules; metal chelates; antibodies; single- or double-stranded nucleic acids; proteins, peptides, amino acids; carbohydrates; lipopolysaccharides; sugar amino acid hybrids; molecules from phage display libraries; biotin; avidin; streptavidin; and combinations thereof.

66. The computer program product, wherein an assay used to measure levels of data elements in training data sets from which candidate biomarkers are identified is different from an assay used to measure data elements in a validation data set used to validate the candidate biomarker.

67. The computer program product, wherein the assay used to measure levels of data elements in training data sets is SELDI.

68. The computer program product of claim 66 or 67, wherein the assay used to measure levels of data elements in validation data sets is an immunoassay.

69. The computer program product of claim 36, wherein the independent discovery data sets are collected from different locations, using different collection protocols, and/or are collected from different populations.

70. The computer program product of claim 36, wherein each discovery data set is from a different clinical trial site.

71. A system comprising:

one or more processors for

(a) receiving input data relating to at least first and second independent discovery data sets wherein:

(i) the data sets comprise a plurality of forms of biological state classes;

(ii) each data set comprises a plurality of data points, wherein each data point exhibits one form of a biological state class and each data set comprises a plurality of data points belonging to each of the classes; and

(iii) each data point comprises a plurality of data elements, each data element characterized by a value, wherein all data points share a plurality of common data elements;

(b) executing computer readable program code providing instructions for qualifying each common data element, independently for each data set, based on the ability of the data element to classify a data point into a biological state class, as a function of data element value and for selecting an initial subset of data elements within each data set; and

(c) executing computer readable program code providing instructions for selecting an intersection subset of data elements from the initial subsets, wherein each data element in the intersection subset is a member of a majority of the initial subsets.

72. The system of claim 71, further comprising one or more devices for providing input data to the one or more processors.

73. The system of claim 72, wherein the one or more devices for providing input data comprises a detector for detecting a characteristic of a data element.

74. The system of claim 73, wherein the detector comprises a mass spectrometer.

75. The system of claim 73, wherein the detector comprises a gene chip reader.

76. The system of claim 71, further comprising a memory for storing a data set of ranked data elements.

77. The system of claim 71, further comprising a database of ranked data elements.

78. The system of claim 71, wherein selecting the initial subsets comprises using the discovery data sets to train a learning algorithm wherein the learning algorithm ranks the data elements based on a quantitative measure of ability to classify.

79. The system of claim 78, wherein the learning algorithm is a supervised learning algorithm.

80. The system of claim 78, wherein the learning algorithm is an unsupervised learning algorithm.

81. The system of claim 78, wherein training comprises support vector machine analysis.

82. The system of claim 78, wherein training comprises linear discrimination analysis.

83. The system of claim 78, wherein training comprises combining support vector machine analysis and linear discrimination analysis.

84. The system of claim 78, wherein training comprises performing unified maximum separability analysis (UMSA).

85. The system of claim 71, wherein the system further executes program code for independently re-sampling data elements in each data set.

86. The system of claim 78, wherein the system further executes program code for selecting candidate biomarkers based on ranking by the learning algorithm and for testing one or more of the candidate biomarkers on a validation data set.

87. The system of claim 71, wherein the biological state class comprises a cell state.

88. The system of claim 71, wherein the biological state class comprises a patient status.

89. The system of claim 71, wherein the biological state class is selected from the group consisting of: presence of a disease; absence of a disease; progression of a disease; risk for a disease; stage of disease; likelihood of recurrence of disease; a genotype; a phenotype; exposure to an agent or condition; a demographic characteristic; resistance to agent, sensitivity to an agent, and combinations thereof.

90. The system of claim 89, wherein the genotype is selected from the group consisting of an HLA haplotype; a mutation in a gene; a modification of a gene, and combinations thereof.

91. The system of claim 89, wherein the agent is selected from the group consisting of a toxic substance, a potentially toxic substance, an environmental pollutant, a candidate drug, and a known drug.

92. The system of claim 89, wherein the demographic characteristic is selected from the group consisting of: age, gender, weight; family history; and history of preexisting conditions.

93. The system of claim 89, wherein sensitivity to an agent comprises responsiveness to a drug.

94. The system of claim 86, wherein the one or more candidate biomarkers are diagnostic of the presence of a disease, risk of developing a disease, risk of recurrence of a disease, or stage of the disease.

95. The system of claim 71, wherein values of the data elements in a data point represent levels and/or frequency of components in a data point sample.

96. The system of claim 95, wherein components are selected from the group consisting of: nucleic acids, proteins, polypeptides, peptides, carbohydrates and modified or processed forms thereof.

97. The system of claim 95, wherein levels of components are measured by an expression profiling assay.

98. The system of claim 97, wherein the expression profiling assay comprises measuring the amount and/or form of a nucleic acid.

99. The system of claim 98, wherein expression profiling comprises measuring amplification, mutation, and/or modification of DNA.

100. The system of claim 97, wherein the expression profiling assay comprises measuring the amount and/or form of a protein, polypeptide or peptide.

101. The system of claim 97, wherein the expression profiling assay comprises mass spectrometry.

102. The system of claim 101, wherein the expression profiling assay comprises SELDI analysis.

103. The system of claim 97, wherein the expression profiling assay comprises measuring the amount and/or form of a carbohydrate.

104. The system of claim 71, wherein data elements of data points comprise data relating to the cellular localization of components in a sample.

105. The system of claim 97, wherein expression profiling comprises:

(a) contacting samples with a substrate comprising binding partners for specifically binding to sample components having selected characteristics and

(b) identifying sample components bound to the substrate.

106. The system of claim 105, wherein binding partners are selected from the group consisting of cationic molecules; anionic molecules; metal chelates; antibodies; single- or double-stranded nucleic acids; proteins, peptides, amino acids; carbohydrates; lipopolysaccharides; sugar amino acid hybrids; molecules from phage display libraries; biotin; avidin; streptavidin; and combinations thereof.

107. The system of claim 71, wherein an assay used to measure levels of data elements in training data sets from which candidate biomarkers are identified is different from an assay used to measure data elements in a validation data set used to validate the candidate biomarker.

108. The system of claim 107, wherein the assay used to measure levels of data elements in training data sets is SELDI.

109. The system of claim 107 or 108, wherein the assay used to measure levels of data elements in validation data sets is an immunoassay.

110. The system of claim 71, wherein the independent discovery data sets are collected from different locations, using different collection protocols, and/or are collected from different populations.

111. The system of claim 71, wherein each discovery data set is from a different clinical trial site.